

Klasifikasi Anak Putus Sekolah di Provinsi Jawa Timur Tahun 2012 Menggunakan Regresi Logistik Biner dan Kohonen *Learning Vector Quantization* (LVQ)

Avinia Aisha Widhesaputri dan Brodjol Sutijo S.U.
Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail: brodjol_su@statistika.its.ac.id

Abstrak—Wajib belajar sembilan tahun merupakan salah satu program pemerintah sebagai solusi untuk meningkatkan angka partisipasi sekolah di Indonesia. Program ini menargetkan angka partisipasi sekolah minimal 95 persen pada akhir tahun 2008. Salah satu masalah yang menghambat pencapaian wajib belajar sembilan tahun adalah anak putus sekolah. Berdasarkan data Kementerian Pendidikan Nasional pada 2008, setiap tahunnya terdapat 1,5 juta remaja di Indonesia yang tidak dapat melanjutkan sekolah. Usaha untuk menyelesaikan masalah tersebut adalah dengan mengidentifikasi siswa putus sekolah, kemudian membantu mereka agar dapat bersekolah kembali serta memberi dukungan sehingga berhasil menyelesaikan wajib belajar 9 tahun. Selanjutnya dilakukan pengelompokan (klasifikasi) anak putus sekolah untuk mengetahui sebaran dan karakteristiknya. Klasifikasi dilakukan dengan model regresi logistik biner dan *Learning Vector Quantization* (LVQ) dengan menggunakan variabel prediktor jenis kelamin, status perkawinan dan status bekerja siswa, tingkat pendidikan dan jenis kelamin kepala rumah tangga, serta pengeluaran, jumlah anggota dan daerah tempat tinggal keluarga. Data merupakan hasil SUSENAS tahun 2012 di Provinsi Jawa Timur. Hasil identifikasi anak putus sekolah dengan model regresi logistik biner mendapatkan ketepatan klasifikasi sebesar 89,6 persen. Sedangkan identifikasi anak putus sekolah dengan jaringan LVQ menggunakan 4 *hidden neuron* menghasilkan ketepatan klasifikasi sebesar 88,9 persen.

Kata Kunci— anak putus sekolah, klasifikasi, *learning vector quantization*, regresi logistik biner

I. PENDAHULUAN

PENDIDIKAN adalah usaha sadar dan terencana untuk mewujudkan suasana dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi diri untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya dalam masyarakat (UU SISDIKNAS no. 20 tahun 2003). Pendidikan merupakan salah satu aspek penting bagi pembangunan bangsa sebagai wadah untuk membentuk generasi yang dibutuhkan di masa mendatang, sehingga hampir semua bangsa menempatkan pendidikan sebagai prioritas utama dalam program pembangunan nasional [1]. Untuk itu, idealnya setiap anak mengikuti pendidikan di sekolah sehingga mereka menjadi lebih matang secara kognitif, afektif maupun motorik. Kenyataannya, tidak semua anak mempunyai kesempatan untuk berpartisipasi dalam pendidikan di sekolah.

Pemerintah mempunyai tanggung jawab dalam menjamin pemerataan kesempatan untuk memperoleh pendidikan bagi semua anak usia sekolah di Indonesia. Penuntasan wajib belajar sembilan tahun merupakan upaya untuk meningkatkan angka partisipasi murni anak usia wajib belajar sembi-

lan tahun hingga mencapai minimal 95 persen pada akhir tahun 2008 [2]. Kenyataannya program ini tidak berjalan sesuai harapan karena sampai saat ini masih banyak anak-anak yang belum menuntaskan pendidikan dasar sembilan tahun [3]. Salah satu masalah dalam pencapaian wajib belajar sembilan tahun adalah siswa yang putus sekolah. Berdasarkan data Kementerian Pendidikan Nasional pada 2008, setiap tahunnya 1,5 juta remaja Indonesia tidak dapat melanjutkan sekolah. Data yang dikumpulkan Kementerian Pemberdayaan Perempuan dan Perlindungan anak pada 2009 menunjukkan bahwa untuk tingkat SMP kasus putus sekolah mencapai rata-rata 150.000 hingga 200.000 anak per tahun. Data tersebut menunjukkan bahwa jumlah anak putus sekolah di Indonesia tergolong tinggi [4].

Usaha untuk menyelesaikan masalah putus sekolah salah satunya adalah dengan mengidentifikasi siswa yang putus sekolah, kemudian membantu mereka supaya dapat bersekolah lagi serta memberi dukungan hingga mereka berhasil lulus SMP. Kegiatan yang berkaitan dengan upaya-upaya mencari dan mengenali (identifikasi) siswa putus sekolah, serta mencari faktor penyebab putus sekolah perlu dilakukan agar penyelesaian masalah tersebut tepat sasaran [2]. Selanjutnya, untuk mengetahui sebaran dan karakteristik anak putus sekolah perlu dilakukan pengelompokan / klasifikasi anak putus sekolah sesuai dengan karakteristiknya.

Berbagai penelitian telah dilakukan untuk menjelaskan masalah putus sekolah, baik ditinjau dari sisi individu maupun dari sisi eksternal. Sugiyanto (1996) menerapkan regresi logistik untuk menganalisis pengaruh faktor sosial, ekonomi dan demografi rumah tangga dalam pemilihan jenis pendidikan jenjang menengah di Pulau Jawa. Supriyadi (2006) meneliti tentang faktor-faktor demografi seperti pendidikan ibu, pengeluaran rumah tangga dan jenis kelamin anak. Berbagai penelitian ini menggunakan faktor ekonomi, sosial, dan demografi anak dan rumah tangga sebagai faktor yang mempengaruhi partisipasi sekolah anak.

Regresi logistik biner merupakan metode regresi yang digunakan untuk mencari hubungan antara variabel respon yang bersifat dikotomis (bersifat nominal atau ordinal dengan dua kategori) dengan satu atau lebih prediktor [5]. Metode ini sesuai untuk melakukan identifikasi anak putus sekolah yang memiliki variabel respon dari dua kategori (putus sekolah dan tidak putus sekolah). Selanjutnya dilakukan perbandingan hasil klasifikasi antara metode klasifikasi klasik dan metode klasifikasi yang telah dikembangkan (*artificial intelligent*). Metode klasifikasi *Neural Network Learning Vector Quantization* dipilih karena menghasilkan output yang cukup sederhana sehingga mudah dalam interpretasi.

II. TINJAUAN PUSTAKA

A. Regresi Logistik Biner

Regresi logistik biner merupakan suatu metode yang digunakan untuk mencari hubungan variabel respon yang bersifat dikotomis dengan satu atau lebih prediktor [5]. Variabel prediktor mengikuti distribusi Bernoulli dengan fungsi probabilitas $P(Y = 1|x) = \pi(x)$.

Regresi logistik menggambarkan probabilitas terjadinya suatu kejadian yang besarnya antara 0 dan 1. Bentuk persamaan regresi logistik adalah sebagai berikut.

$$g(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (1)$$

Dengan probabilitas kejadian dihitung dari persamaan berikut.

$$\pi(x) = \frac{\exp(g(x))}{1 + \exp(g(x))} \quad (2)$$

Parameter dalam model regresi logistik biner diduga menggunakan metode *Maximum Likelihood Estimator* (MLE) [5], yang akan memberikan pendekatan estimasi parameter pada model regresi logistik dimana fungsi *likelihood* adalah fungsi peluang pengamatan. Hasil parameter β didapatkan dengan memaksimumkan fungsi *likelihood* berikut.

$$L(\beta) = \ln [l(\beta)] = \sum_{i=1}^n \{y_i \ln [\pi(x_i)] + (1 - y_i) \ln [1 - \pi(x_i)]\} \quad (3)$$

Nilai maksimum dari persamaan (3) secara matematis akan lebih mudah didapatkan jika didapatkan dari nilai \ln *likelihood*. Nilai taksiran parameter β yang diturunkan dari fungsi \ln *likelihood* dapat diperoleh menggunakan metode *Newton-Raphson* [5].

Pengujian signifikansi parameter model regresi logistik biner dilakukan secara individu maupun serentak. Pengujian secara individu dilakukan dengan statistik uji Wald [6]. Hipotesis yang digunakan dalam pengujian signifikansi parameter secara individu adalah sebagai berikut.

$$H_0 : \beta_k = 0 \quad H_1 : \beta_k \neq 0$$

Statistik uji :

$$W^2 = \frac{\hat{\beta}_k^2}{se(\hat{\beta}_k^2)} \quad (4)$$

Sementara pengujian signifikansi parameter secara serentak dilakukan dengan statistik uji G [6]. Dengan hipotesis sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{minimal ada satu } \beta_j \neq 0, \text{ dimana } j = 1, 2, \dots, p$$

Statistik uji :

$$G = -2 \ln \left[\frac{\binom{n_1}{n} \binom{n_0}{n}^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{1 - y_i}} \right] \quad (5)$$

Daerah penolakan H_0 untuk uji signifikansi parameter adalah jika $p_{value} < \alpha$.

Setelah didapatkan model regresi logistik biner yang terbaik, dilakukan pengujian untuk mengetahui kesesuaian hasil prediksi model dengan hasil pengamatan. Pengujian ini dilakukan menggunakan statistik uji *Chi-square* [6]. Hipotesis yang dilakukan dalam pengujian ini adalah sebagai berikut.

H_0 : model sesuai (tidak ada perbedaan antara hasil observasi dengan hasil prediksi model)

H_1 : model tidak sesuai (ada perbedaan antara hasil observasi dengan hasil prediksi model)

Statistik uji :

$$\chi^2 = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} \quad (6)$$

$$\text{dengan } o_k = \sum_{j=1}^{n'_k} y_j; \bar{\pi}_k = \sum_{j=1}^{n'_k} \frac{m_j \hat{\pi}_j}{n'_k}$$

Keterangan :

g = jumlah kelompok

n_k = total pengamatan pada kelompok ke- k

o_k = jumlah nilai respon pada kelompok ke- k

m_j = banyaknya pengamatan yang memiliki nilai $\hat{\pi}_j$

$\bar{\pi}_k$ = rata-rata taksiran probabilitas

Daerah penolakan dari H_0 adalah jika nilai $\chi^2 > \chi^2_{(\alpha, g-2)}$ atau $p_{value} < \alpha$.

Interpretasi koefisien dalam model regresi logistik biner menggunakan *odds ratio* ($\exp(\beta)$) yang merupakan ukuran risiko atau kecenderungan untuk mengalami kejadian tertentu antara satu kategori dengan kategori lainnya, didefinisikan sebagai *ratio* dari *odds* untuk $x_j = 1$ terhadap $x_j = 0$. Untuk prediktor yang berskala kontinyu, interpretasi koefisien β_j pada model regresi logistik adalah setiap kenaikan c unit pada prediktor akan menyebabkan terjadinya $y = 1$ adalah $\exp(c \cdot \beta_j)$ kali lebih besar [5].

B. Learning Vector Quantization

Jaringan *learning vector quantization* (LVQ) merupakan jaringan Kohonen yang telah ditransformasikan agar dapat menggunakan metode *supervised learning*. Metode *learning* disebut *supervised* jika target telah diketahui sebelumnya, karena algoritma dalam *supervised learning* memanfaatkan informasi dari setiap anggota kelas untuk mendeteksi perbedaan antara pola output hasil *learning* dengan pola target [7]. Dalam LVQ input akan bersaing pada suatu *competitive layer* untuk dapat masuk dalam suatu kelas. Hasil dari *competitive layer* ini berupa subkelas, yang kemudian akan dihubungkan menjadi target kelas oleh fungsi aktivasi [8].

Algoritma jaringan LVQ untuk melakukan klasifikasi adalah sebagai berikut [11].

- 1) Inisialisasi vektor bobot (w_j) yang mewakili kelas dan *learning rate* (α).
- 2) Melakukan langkah 3 sampai 7 jika kondisi berhenti tidak terpenuhi.
- 3) Melakukan langkah 4 sampai 5 untuk setiap vektor input.
- 4) Menentukan nilai J sedemikian agar D_j minimum,

$$D_j = \sqrt{\sum_{i=1}^n (x_i - w_{ij})^2} \quad (7)$$

- 5) Memperbarui nilai bobot w_j di sekitar vektor input pemenang, dengan persamaan:

Jika $C_j = T$ maka

$$w_j (\text{baru}) = w_j (\text{lama}) + \alpha [x - w_j (\text{lama})] \quad (8)$$

Jika $C_j \neq T$ maka

$$w_j (\text{baru}) = w_j (\text{lama}) + \alpha [x - w_j (\text{lama})] \quad (9)$$

Keterangan:

C_j = Kelas yang diwakili neuron ke- j (hasil *learning*)

T = Kelas target untuk input

$w_j (\text{baru})$ = bobot w_j yang baru

$w_j (\text{lama})$ = bobot w_j awal

x = vektor *input* = $(x_1, x_2, x_3, \dots, x_n)$

6) Mereduksi nilai *learning rate* (α).

7) Menguji kondisi berhenti.

Jumlah subkelas selalu lebih besar dibandingkan jumlah target kelas yang telah ditentukan [8]. Fungsi aktivasi *linear* akan menghubungkan subkelas dari vektor input pemenang dengan kelas target.

C. Ukuran Ketepatan Klasifikasi

Evaluasi prosedur klasifikasi adalah suatu evaluasi yang melihat peluang kesalahan klasifikasi yang dilakukan oleh suatu fungsi klasifikasi [10]. Ukuran yang digunakan adalah *classification accuracy*, *sensitivity*, dan *specificity* yang dapat ditentukan menggunakan nilai dalam *confusion matrix* seperti yang ditunjukkan pada Tabel 1. *Sensitivity* merupakan ukuran ketepatan dari suatu kejadian yang diinginkan. Sedangkan *specificity* adalah suatu ukuran yang menyatakan persentase kejadian-kejadian yang tidak diinginkan.

Tabel 1. Confusion Matrix

Kenyataan	Prediksi	
	Kelas 0	Kelas 1
Kelas 0	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
Kelas 1	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)

Keterangan :

TP = jumlah observasi dari kelas 0 yang tepat diprediksikan sebagai kelas 0

TN = jumlah observasi dari kelas 1 yang tepat diprediksikan sebagai kelas 1

FP = jumlah observasi dari kelas 1 yang salah diprediksi sebagai kelas 0

FN = jumlah observasi dari kelas 0 yang salah diprediksi sebagai kelas 1

Ukuran ketepatan klasifikasi dihitung dengan persamaan berikut.

$$Accuracy(\%) = \frac{TP+TN}{TP+FP+FN+TN} \quad (10)$$

$$Sensitivity(\%) = \frac{TP}{TP+FN} \quad (11)$$

$$Specificity(\%) = \frac{TN}{FP+TN} \quad (12)$$

D. Anak Putus Sekolah

Putus sekolah adalah proses berhentinya siswa secara terpaksa dari suatu lembaga pendidikan tempat dia belajar atau terlantarnya anak dari sebuah lembaga pendidikan formal yang disebabkan oleh berbagai faktor. Wajib belajar merupakan salah satu program yang gencar digalakkan oleh Departemen Pendidikan Nasional [1]. Program ini mewajibkan setiap warga negara Indonesia untuk bersekolah selama sembilan tahun pada jenjang pendidikan dasar, yaitu dari tingkat kelas 1 Sekolah Dasar (SD/MI) hingga kelas 9 Sekolah Menengah Pertama (SMP/MTs).

Terdapat beberapa faktor yang mempengaruhi anak putus sekolah yang dapat dibagi menjadi faktor internal dan eksternal. Faktor internal merupakan faktor yang datangnya berasal dari anak itu sendiri, misalnya anak yang malas berangkat sekolah karena tidak memiliki minat dalam pendidikan. Sedangkan faktor eksternal yang mempengaruhi anak putus sekolah antara lain adalah faktor ekonomi, demografi, sosial budaya dan geografi [4]. Besarnya biaya yang harus dikeluarkan untuk menempuh pendidikan di Indonesia menjadikan faktor ekonomi sebagai faktor utama

yang menyebabkan anak putus sekolah. Sedangkan faktor demografi dan sosial budaya mempengaruhi keputusan orang tua dalam menentukan kepentingan pendidikan bagi anak-anaknya. Faktor geografi yang mempengaruhi putus sekolah misalnya jarak rumah dengan sarana sekolah terdekat atau mudahnya akses anak ke sekolah.

III. METODOLOGI PENELITIAN

A. Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder angka lama sekolah yang bersumber dari data SUSENAS tahun 2012 di Provinsi Jawa Timur. Dari seluruh data diambil data rumah tangga yang memiliki anak berusia 16-18 tahun berjumlah 3.794 rumah tangga.

Variabel penelitian yang digunakan dalam penelitian ini diberikan pada Tabel 2.

Tabel 2. Variabel

Variabel	Nama Variabel	Skala Data	Kategori
Y	Putus Sekolah	Nominal	1 = Tidak Putus Sekolah 0 = Putus Sekolah
X ₁	Jenis Kelamin	Nominal	1 = Laki-Laki 0 = Perempuan
X ₂	Status Perkawinan	Nominal	1 = Belum Kawin 0 = Kawin
X ₃	Status Bekerja	Nominal	1 = Tidak Bekerja 0 = Bekerja
X ₄	Tingkat Pendidikan Kepala Rumah Tangga	Ordinal	0 = Tidak Tamat SD 1 = Tamat SD 2 = Tamat SMP 3 = Tamat SMA 4 = Tamat PT
X ₅	Jenis Kelamin Kepala Rumah Tangga	Nominal	1 = Laki-Laki 0 = Perempuan
X ₆	Rata-Rata Pengeluaran Rumah Tangga per Kapita	Rasio	
X ₇	Jumlah Anggota Rumah Tangga	Rasio	
X ₈	Lokasi Tempat Tinggal	Nominal	1 = Kota 0 = Desa

B. Langkah Penelitian

Langkah analisis yang digunakan dalam penelitian ini adalah sebagai berikut.

- 1) Membagi data untuk *training* dan *testing* secara acak dengan pembagian 80:20.
- 2) Melakukan pemodelan regresi logistik biner, meliputi tahapan berikut.
 - a) Menguji dependensi setiap variabel prediktor dengan variabel respon.
 - b) Melakukan pengujian signifikansi parameter model regresi logistik biner secara individu dan serentak.
 - c) Membentuk model regresi logistik biner dengan variabel prediktor yang signifikan.
 - d) Menguji kesesuaian model yang terbentuk.
 - e) Menghitung ketepatan klasifikasi model.
- 3) Melakukan klasifikasi anak putus sekolah dengan metode LVQ menggunakan jumlah *hidden* neuron 4,6,8, dan 10, meliputi tahapan berikut.
 - a) Menentukan nilai awal vektor bobot dan *learning rate*.
 - b) Menentukan nilai J yang meminimumkan D_j .

- c) Memperbarui nilai vektor bobot di sekitar vektor pemenang.
 - d) Menguji kondisi penghentian iterasi.
 - e) Menghitung ketepatan klasifikasi jaringan.
 - f) Menentukan jumlah *hidden* neuron terbaik berdasarkan performansi jaringan.
- 4) Membandingkan ketepatan hasil klasifikasi kedua metode.
 - 5) Menarik kesimpulan.

IV. ANALISIS DAN PEMBAHASAN

A. Karakteristik Anak Putus Sekolah di Jawa Timur

Pada penelitian ini dilakukan analisis deskriptif untuk mengetahui perbedaan karakteristik antara anak yang mengalami putus sekolah maupun tidak. Berdasarkan data SUSENAS tahun 2012 di Provinsi Jawa Timur, didapatkan 3.794 data yang terdiri dari 450 anak putus sekolah dan 3.344 anak tidak putus sekolah.

Tabel 3. Tabulasi Silang Setiap Variabel Prediktor

Variabel	Kategori	Putus Sekolah (%)	Tidak Putus Sekolah
Jenis Kelamin	Laki-Laki	253 (12,6)	1747 (87,4)
	Perempuan	197 (11,0)	1597 (89,0)
Status Perkawinan	Belum Kawin	385 (10,6)	3239 (89,4)
	Kawin	65 (38,2)	105 (61,8)
Status Bekerja	Tidak Bekerja	253 (8,3)	2799 (91,7)
	Bekerja	197 (26,5)	545 (73,5)
Tingkat Pendidikan Kepala Rumah Tangga	Tidak Tamat SD	250 (24,8)	759 (75,2)
	Tamat SD	148 (11,4)	1145 (88,6)
	Tamat SMP	28 (5,7)	462 (94,3)
	Tamat SMA	17 (2,4)	697 (97,6)
Jenis Kelamin Kepala Rumah Tangga	Laki-Laki	403 (11,8)	3007 (88,2)
	Perempuan	47 (12,2)	337 (87,8)
Daerah Tempat Tinggal	Kota	154 (7,2)	1979 (92,8)
	Desa	296 (17,8)	1365 (82,2)

Tabel 3 menunjukkan karakteristik data anak yang mengalami putus sekolah maupun tidak yang disajikan dalam tabulasi silang. Sehingga dapat disimpulkan bahwa mayoritas anak yang mengalami putus sekolah di Jawa Timur memiliki karakteristik berjenis kelamin laki-laki, sudah kawin, sudah bekerja, kepala rumah tangga tidak tamat SD dan berjenis kelamin perempuan, serta bertempat tinggal di daerah desa.

B. Model Regresi Logistik Biner

Pembentukan regresi logistik biner dilakukan dengan data *training* sebanyak 3.035 pengamatan. Sebelum membentuk model, dilakukan uji dependensi masing-masing variabel prediktor dengan respon. Uji dependensi menunjukkan bahwa variabel jenis kelamin anak (X_1) dan kepala rumah tangga (X_5) saling independen dengan variabel respon pada $\alpha = 0,05$. Sehingga kedua variabel tersebut tidak disertakan dalam pembentukan model regresi logistik biner.

Tabel 4. Hasil Uji Independensi

Variabel	$\chi^2_{(hit)}$	P _{value}
Jenis Kelamin	2,520	0,062 ^{*)}
Status Perkawinan	118,424	0,000
Status Bekerja	190,379	0,000
Tingkat Pendidikan Kepala Rumah Tangga	268,817	0,000
Jenis Kelamin Kepala Rumah Tangga	0,059	0,430 ^{*)}
Lokasi Tempat Tinggal	100,380	0,000

Keterangan : ^{*)} tidak signifikan pada $\alpha = 0,05$

Pengujian individu dilakukan untuk melihat signifikansi parameter variabel yang akan digunakan dalam model. Hasil uji individu untuk setiap variabel prediktor menunjukkan bahwa semua variabel prediktor yang disertakan dalam pembentukan model regresi logistik biner memiliki pengaruh signifikan terhadap model pada $\alpha = 0,05$. Sementara berdasarkan pengujian parameter secara serentak disimpulkan bahwa semua variabel memiliki pengaruh signifikan secara bersama-sama dengan nilai statistik uji *Chi-square* sebesar 370,839.

Tabel 5. Parameter Model Regresi Logistik Biner

	B	Wald	P _{value}	Odds Ratio
Status Perkawinan	1,132	28,94224	0,000	3,104
Status Bekerja	1,017	65,13721	0,000	2,767
Tingkat Pendidikan Kepala RT		73,59615	0,000	
Tidak Tamat SD	- 0,773	33,73224	0,000	0,462
Tamat SD	- 1,247	28,38085	0,000	0,287
Tamat SMP	- 1,922	39,78919	0,000	0,146
Tamat SMA	- 1,090	6,591325	0,010	0,336
Pengeluaran RT per Kapita	- 1,370	13,99032	0,000	0,254
Jumlah Anggota RT	0,123	8,661001	0,003	1,131
Lokasi Tempat Tinggal	0,342	6,849332	0,009	1,408
Konstan	- 1,876	39,45797	0,000	0,153

Berdasarkan koefisien parameter yang ditunjukkan pada Tabel 5, didapatkan model regresi logistik biner dengan persamaan sebagai berikut.

$$g(x) = 0,47 + 1,13X_2(1) + 1,02X_3(1) - 1,09X_4(1) - 0,32X_4(2) + 0,16X_4(3) + 0,83X_4(4) + 1,37X_6 - 0,12X_7 + 0,34X_8(1)$$

Dari persamaan logit yang terbentuk, dapat dihitung nilai probabilitas seorang anak mengalami putus sekolah atau tidak berdasarkan karakteristik faktornya seperti berikut.

$$g(x) = 0,47 + 1,13(0) + 1,02(0) - 1,09(1) - 0,32(0) + 0,16(0) + 0,83(0) + 1,37(0,341) - 0,12(5) + 0,34(0) = -0,765$$

$$\pi(x) = \frac{\exp(-0,765)}{1 + \exp(-0,765)} = 0,318$$

Nilai probabilitas dari karakteristik mayoritas anak putus sekolah di Jawa Timur (anak yang sudah kawin dan bekerja, pendidikan kepala rumah tangga tidak tamat SD, pengeluaran rumah tangga per kapita 0,341 juta, beranggota keluarga 5 dan tinggal di daerah desa) adalah 0,318. Jika dibandingkan dengan nilai *cut-off* sebesar 0,5 maka anak dengan karakteristik tersebut akan diidentifikasi sebagai anak yang putus sekolah.

$$g(x) = 0,47 + 1,13(1) + 1,02(1) - 1,09(0) - 0,32(0) + 0,16(0) + 0,83(1) + 1,37(0,541) - 0,12(4) + 0,34(1) = 4,044$$

$$\pi(x) = \frac{\exp(4,044)}{1 + \exp(4,044)} = 0,983$$

Sedangkan nilai probabilitas dari karakteristik mayoritas anak yang tidak putus sekolah di Jawa Timur (anak yang belum kawin maupun bekerja, tingkat pendidikan kepala rumah tangga tamat SMA, pengeluaran rumah tangga per kapita 0,541 juta, beranggota keluarga 4 dan tinggal di daerah kota) adalah 0,983 yang berarti akan masuk ke kelas anak tidak putus sekolah.

Odds ratio merupakan kecenderungan untuk mengalami kejadian tertentu antara satu kategori dengan kategori lainnya. Misalnya anak yang sudah menikah memiliki ke-

cenderung mengalami putus sekolah sebanyak 3 kali lebih tinggi dibanding anak yang belum menikah. Anak yang kepala keluarganya tamat SD cenderung mengalami putus sekolah 0,462 kali lebih tinggi dibanding anak yang kepala keluarganya tidak tamat SD, atau anak yang kepala keluarganya tamat SMP cenderung tidak mengalami putus sekolah 4 kali lebih tinggi dibanding anak yang kepala keluarganya tidak tamat SD. Dan setiap kenaikan pengeluaran per kapita keluarga sebesar satu juta, kecenderungan anak untuk tidak putus sekolah meningkat 3 kali lebih tinggi.

Setelah didapatkan model regresi logistik biner terbaik, dilakukan uji kesesuaian model untuk mengetahui apakah hasil prediksi model telah sesuai dengan hasil observasi. Berdasarkan hasil uji *Hosmer-Lemeshow* dapat disimpulkan bahwa model yang didapatkan telah sesuai, dengan p value sebesar 0,308. Ketepatan klasifikasi dari model didapatkan dari perbandingan hasil prediksi dengan hasil observasi data *testing* yang diberikan pada Tabel 6.

Tabel 6. Confusion Matrix Klasifikasi Regresi Logistik

Kenyataan	Prediksi		Persentase Benar
	Putus Sekolah	Tidak Putus Sekolah	
Putus Sekolah	8	75	9,6%
Tidak Putus Sekolah	4	672	99,4%
Persentase Keseluruhan			89,6%

Model regresi logistik biner yang terbentuk dapat mengidentifikasi anak tidak putus sekolah dengan nilai ketepatan mencapai 99,4%. Namun hanya dapat mengidentifikasi anak putus sekolah dengan tepat sebesar 9,6%. Secara keseluruhan model regresi logistik biner yang terbentuk dapat melakukan identifikasi anak putus sekolah sebesar 89,6%.

C. Learning Vector Quantization

Sebuah jaringan LVQ dibuat dengan input berupa matriks berukuran 8×2 yang berisi nilai minimum dan maksimum masing-masing variabel prediktor. *Learning rate* yang digunakan sebesar 0,01 dengan metode pembelajaran LVQ1 dan proporsi setiap kelas target [0,5 0,5]. Proses *learning* dengan beberapa jumlah *hidden* neuron (node) dilakukan pada *layer* pertama untuk mendapatkan jumlah node yang optimal.

Jumlah subkelas harus lebih banyak dibandingkan jumlah kelas target yang sebanyak 2. Dipilih jumlah node genap agar memudahkan pembagian proporsi setiap kelas. Pada penelitian ini digunakan jumlah node sebanyak 4, 6, 8 dan 10 sebagai perbandingan untuk mengetahui apakah ada pengaruh jumlah node yang digunakan terhadap performansi klasifikasi. Proses *learning* diulang sebanyak 10 kali untuk setiap jumlah node, kemudian dipilih jaringan hasil *learning* terbaik berdasarkan *Mean Squared Error* (mse) jaringan.

Berikut adalah perbandingan jaringan hasil *learning* terbaik untuk setiap node yang digunakan. Ukuran yang diamati adalah nilai performansi (mse), akurasi data *training*, dan akurasi data *testing*.

Tabel 7. Perbandingan Hasil *Learning* Setiap Node

node	Performansi	Akurasi Training	Akurasi Testing
4	0,1219	0,8781	0,8893
6	0,1209	0,8791	0,8906
8	0,1213	0,8787	0,8906
10	0,1216	0,8784	0,8906

Tabel 7 menunjukkan bahwa performansi klasifikasi terbaik dihasilkan jaringan LVQ yang menggunakan 4 node. Bobot *layer* pertama dari jaringan LVQ tersebut diberikan pada Tabel 8.

Tabel 8. Bobot Jaringan LVQ

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Subkelas 1	1,054	2,897	3,769	-0,168	3,795	3,600	8,586	2,760
Subkelas 2	2,975	1,889	3,395	-1,351	1,800	2,041	11,364	3,252
Subkelas 3	1,470	1,022	1,146	1,894	1,092	0,448	4,190	1,545
Subkelas 4	1,429	1,002	1,090	3,985	1,085	0,723	4,499	1,147

Selanjutnya dihitung ukuran ketepatan klasifikasi dari jaringan LVQ terpilih, yaitu jaringan dengan 4 node, untuk melakukan klasifikasi terhadap data *testing*. Hasil ketepatan klasifikasi untuk jaringan LVQ diberikan pada Tabel 9.

Tabel 9. Confusion Matrix Klasifikasi Jaringan LVQ

Kenyataan	Prediksi		Persentase Benar
	Putus Sekolah	Tidak Putus Sekolah	
Putus Sekolah	0	83	0%
Tidak Putus Sekolah	1	675	99,8%
Persentase Keseluruhan			88,9%

Jaringan LVQ dengan 4 *hidden* neuron yang dihasilkan hanya mampu mengidentifikasi anak putus sekolah dengan benar. Namun mampu mengidentifikasi anak yang tidak putus sekolah dengan tepat sebesar 99,8%. Secara keseluruhan, jaringan LVQ yang dibentuk dapat melakukan identifikasi anak putus sekolah dengan nilai ketepatan sebesar 88,9%.

D. Perbandingan Hasil Klasifikasi

Setelah didapatkan hasil klasifikasi menggunakan model regresi logistik biner dan jaringan LVQ, dilakukan perbandingan ketepatan klasifikasi kedua metode tersebut. Perbandingan ini dilakukan untuk mengetahui metode klasifikasi yang lebih sesuai dalam melakukan identifikasi anak putus sekolah.

Tabel 10. Perbandingan Ketepatan Klasifikasi Metode

Ukuran	Regresi Logistik Biner		Learning Vector Quantization	
	Training	Testing	Training	Testing
Accuracy	89,2%	89,6%	87,8%	88,9%
Sensitivity	7,9%	9,6%	1,4%	-
Specificity	99,3%	99,4%	99,7%	99,8%

Dilihat dari nilai akurasinya, dapat disimpulkan bahwa model regresi logistik biner lebih sesuai digunakan untuk melakukan identifikasi anak putus sekolah. Hal ini didukung dengan nilai sensitivitas metode regresi logistik biner yang juga lebih besar.

V. KESIMPULAN DAN SARAN

Berdasarkan hasil klasifikasi anak putus sekolah menggunakan regresi logistik biner dan *Learning Vector Quantization* dapat diambil kesimpulan sebagai berikut.

- 1) Didapatkan model regresi logistik biner untuk identifikasi anak putus sekolah menggunakan variabel prediktor status kawin dan bekerja anak, tingkat pendidikan orang tua, serta rata-rata pengeluaran, jumlah anggota, dan lokasi tempat tinggal rumah tangga. Model logit yang terbentuk adalah $g(x) = -1,876 + 1,132X_2(1) + 1,017X_3(1) - 0,77X_4(1) - 1,25X_4(2) - 1,92X_4(3) - 1,09X_4(4) - 1,37X_6 +$

$0,12X_7 + 0,34X_8(1)$. Peluang anak untuk mengalami putus sekolah akan meningkat jika anak memiliki karakteristik sudah kawin, sudah bekerja, tingkat pendidikan kepala keluarga rendah, rata-rata pengeluaran rumah tangga rendah, jumlah anggota keluarga besar, atau tinggal di daerah pedesaan. Identifikasi anak putus sekolah dengan model regresi logistik yang terbentuk menghasilkan ketepatan klasifikasi akurasi sebesar 89,6%.

- 2) Jaringan *Learning Vector Quantization* terbaik yang didapatkan menggunakan 4 *hidden* neuron, klasifikasi data menggunakan titik pusat hasil *learning* menghasilkan ketepatan klasifikasi akurasi sebesar 88,9%.
- 3) Model regresi logistik biner lebih sesuai digunakan untuk melakukan identifikasi anak putus sekolah, karena dapat menghasilkan nilai akurasi dan sensitivitas yang lebih besar dibandingkan metode jaringan LVQ.

Hasil klasifikasi yang didapatkan dari kedua metode sama-sama menghasilkan nilai akurasi yang lumayan tinggi. Namun dilihat dari nilai *sensitivity* yang sangat kecil (bahkan 0 pada LVQ), maka terlihat bahwa klasifikasi yang dihasilkan belum baik. Sehingga disarankan untuk penelitian selanjutnya agar digunakan proporsi ukuran data yang seimbang untuk masing-masing kelas.

DAFTAR PUSTAKA

- [1] Firmansyah, Y. (2013). *Peran Unit Rehabilitasi Sosial Karya Mandiri Kabupaten Pemalang dalam Pemerataan Pendidikan Bagi Anak Putus Sekolah*. *Solidarity*, 2, 136-143.
- [2] Sumarmi. (2009). *Pendekatan Bagging CART untuk Klasifikasi Karakteristik Anak Putus Sekolah di Jambi*. Theses, ITS, S2 Statistika, Surabaya.
- [3] Fausett, L. (1994). *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. New York: Prentice Hall.
- [4] Alifianto, A. (2008). Kuliah Kerja Nyata Wajib Belajar 9 Tahun. Diambil kembali dari <http://www.pewarta-kabarindonesia.blogspot.com>
- [5] Agresti, A. (1996). *Categorical Data Analysis*. New York: John Wiley & Sons.
- [6] Hosmer, D. W., & Lemeshow. (2000). *Applied Logistic Regression*. New York: John Wiley & Sons.
- [7] Kusumadewi, Sri, & Hartati, Sri. (2006). *Neuro Fuzzy: Integrasi Sistem Fuzzy & Jaringan Syaraf*. Yogyakarta: Graha Ilmu.
- [8] Demuth, Howard, & Beale, Mark. (2002). *Neural Network Toolbox User's Guide*. Massachusetts: The Math Works, Inc.
- [9] Haykin, S. (1999). *Neural Networks : A Comprehensive Foundation*. New Jersey: Prentice Hall.
- [10] Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis Sixth Edition*. New Jersey: Pearson Education.
- [11] Siang, J. J. (2005). *Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan Matlab*. Yogyakarta: ANDI.
- [12] Anand, R., Mehrotra, K. G., Mohan, C. K., & Ranka, S. (1993). *An Improved Algorithm for Neural Network Classification of Imbalanced Training Sets*. *IEEE Transactions on NN*, 962-969.
- [13] Benediktsson, J., Swain, P., & Ersoy, O. (1990). *Neural Network Approaches Versus Statistical Methods in Classification of Multisource Remote Sensing Data*. *IEEE Transactions in Remote Sensing*, 540-552.
- [14] Ettaouil, M., Abdelatifi, E., Belhabib, F., & El Moutaouakil, K. (2012, November). *Learning Algorithm of Kohonen Network With Selection Phase*. *WSES Transactions on Computers*, 11(11), 387-389.
- [15] Hadi, W. A. (2013). *Pemodelan Faktor-Faktor yang Mempengaruhi Prestasi Mahasiswa Pasca Sarjana ITS dengan Regresi Logistik dan Neural Network*. Tugas Akhir, ITS, S1 Statistika, Surabaya.
- [16] Kohonen, T. (1990). *The Self-Organizing Map*. *Proceedings of the IEEE*, 1464-1480.
- [17] Vesanto, J., & Alhoniemi, E. (2000). *Clustering of the Self-Organizing Map*. *IEEE Transactions on NN*, 586-600.